

On the convergence of the Fitness-Complexity Algorithm

Emanuele Pugliese, Andrea Zaccaria, and Luciano Pietronero

December 15, 2014

Abstract

We investigate the convergence properties of an algorithm which has been recently proposed to measure the competitiveness of countries and the quality of their exported products. These quantities are called respectively Fitness F and Complexity Q . The algorithm was originally based on the adjacency matrix M of the bipartite network connecting countries with the products they export, but can be applied to any bipartite network. The structure of the adjacency matrix turns to be essential to determine which countries and products converge to non zero values of F and Q . Also the speed of convergence to zero depends on the matrix structure. A major role is played by the shape of the ordered matrix and, in particular, only those matrices whose diagonal does not cross the empty part are guaranteed to have non zero values as outputs when the algorithm reaches the fixed point. We prove this result analytically for simplified structures of the matrix, and numerically for real cases. Finally, we propose some practical indications to take into account our results when the algorithm is applied.

1 Introduction

Among the many theories which try to explain the features of the economic development of countries (see for example [Barro and Sala-i Martin, 2004]), a recent, data driven and innovative approach is the one known as *Economic Complexity*. This approach is based on the country-product matrix M , whose binary elements indicates if the given country exports a certain products or not (in particular, if the country shows a Revealed Comparative Advantage [Balassa, 1965] in that market). In contrast with the Ricardian specialization paradigm [Ricardo, 1817], the M matrix is, when suitably ordered, not block diagonal but triangular: this means that ubiquitous products are produced by all countries, while rare products are produced only by diversified countries. In a first attempt to use this information to evaluate the goodness of countries and products, Hidalgo and Hausmann [Hidalgo and Hausmann, 2009] proposed an algorithm on the matrix M . Tacchella and al. [Tacchella et al., 2012] deeply revisited the algorithm and proposed a new metrics to estimate the country's potential of growth, called the *Fitness* F , in terms of the quality, or *Complexity* Q , of its exported products. See [Cristelli et al., 2013] for a comparison between the two approaches.

$$\tilde{F}_c^{(n)} = \sum_p M_{cp} Q_p^{(n-1)} \quad (1)$$

$$F_c^{(n)} = \frac{\tilde{F}_c^{(n)}}{\langle \tilde{F}_c^{(n)} \rangle_c} \quad (3)$$

$$\tilde{Q}_p^{(n)} = \frac{1}{\sum_c M_{cp} \frac{1}{F_c^{(n-1)}}} \quad (2)$$

$$Q_p^{(n)} = \frac{\tilde{Q}_p^{(n)}}{\langle \tilde{Q}_p^{(n)} \rangle_p} \quad (4)$$

where F_c is the Fitness of country c and Q_p is the complexity of product p . All the values are determined only by the Country-Product Matrix, M . We point out that, obviously, in principle it is possible to apply this algorithm to any bipartite network defined by the matrix M , so the definitions of countries and products as the rows and columns of the matrix is somehow reductive. However, since its first and main use so far is this one, in the following we will use the expressions country and product to denote the row and column elements.

In this short note we analyze the characteristics of M required to have convergence of the algorithm to numbers strictly greater than zero, for both every F_c and Q_p . We will first show in section 2 a class of M for which it is possible to solve analytically the algorithm in closed form and find these characteristics; we therefore produce an *ansatz* for the general case. We will then show in section 3 that the insight is numerically supported for any analyzed random matrix and in section 4 we will show how this is valid also for the matrices used in economic complexity. In the conclusions, in section 6, we will explain the effect of this result on the economic complexity field and we will give some practical indications for the use of the algorithm in light of our findings.

Finally there will be an additional appendix, A, in which we will try to generalize our findings to a wider class of algorithms similar to the one we have studied.

2 A theoretical example

2.1 The Matrix Class

We will consider a specific class of matrices. An example of the class is presented in equation 5. The generic matrix in the class is composed of 4 blocks, one of which of zeros. The other three blocks will not be composed only of zeros but they will have some 1s. The density of 1s in the block can be in principle any value between 0 and 1, but to allow for a closed form solution each block has to have the same number of 1s and 0s in all its columns and rows. It is interesting

to note that, since the algorithm holds for reordering of rows and columns, a matrix is also a member of the class if it is possible to reorder its rows and columns in such a way to obtain a matrix in the class.

We will number the four blocks from 1 to 4 anti-clockwise starting from the top right corner. In the following we will always assume that the block 4, the bottom right block, is the block with only zeros. Alternatively, will also refers to the block 4 as the *external area*. We will also call the blocks 1 and 3 the *frontier*, block 2 the *internal area*.

In the following example of this class of matrices the four densities are, anti-clockwise from the top right corner, $1/2$, 1 , $1/3$, 0 .

$$\begin{matrix} & & C_1 & & C_2 \\ R_2 & \left\{ \begin{array}{c|cccccc} 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right. \\ R_1 & \left\{ \begin{array}{c|cccccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right. \end{matrix} \quad (5)$$

The key variables will be the sizes of the four blocks: their height R_1 and R_2 and their width C_1 and C_2 . It will be revealed instead as almost unimportant the density of 1s in the non-empty blocks. To make the notation easier in the following we will assume that all the densities of the three non-empty blocks are equal to 1: the three non empty blocks (1, 2 and 4) will be composed only of 1s. In subsection 2.4 we will explain what changes when we relax this assumption.

2.2 Computations

The symmetry of the matrix is particularly useful because we can state, for symmetry reasons, that all the countries in the R_2 upper rows have the same fitness, and the same is true for the countries in the R_1 lower rows. This is obvious since their rows are equal or can be made equal with a simple rearrangement of the columns. The same is also true for the complexity of the first C_1 products and the complexity of the last C_2 products.

Therefore the fitness at the iteration n of the algorithm can be represented with just two numbers, with an obvious change of notation

$$F^{(n)} = \begin{pmatrix} a \\ b \end{pmatrix} \quad (6)$$

It is easy to compute the next iteration of complexities from equation 2,

$$\tilde{Q}^{(n+1)} = \begin{pmatrix} \frac{ab}{R_2b + R_1a} \\ \frac{a}{R_2} \end{pmatrix} \quad (7)$$

Since we are not interested in the Complexities, to avoid unnecessary computations we will not normalize accordingly to 4. In the next steps we will compute the fitness values and we will normalize them: any multiplicative factor to all the complexities would be removed in any case at that step. Thanks to this property we can rewrite the Complexities in an easier way for the next computation, in particular as

$$\tilde{Q}'^{(n+1)} = \begin{pmatrix} R_2^2b \\ R_2^2b + R_1R_2a \end{pmatrix} \quad (8)$$

where \tilde{Q}' is proportional to \tilde{Q} .

We can now compute the iteration $n+1$ for the fitness from equation 1. Since we are using \tilde{Q}' instead of Q , the values obtained by equation 1 will be proportional to \tilde{F} . Defining these as \tilde{F}' , we obtain

$$\tilde{F}'^{(n+1)} = \begin{pmatrix} C_1R_2^2b + C_2R_2^2b + C_2R_1R_2a \\ C_1R_2^2b \end{pmatrix} = \begin{pmatrix} C_1R_2^2b + C_2R_2^2b + C_2R_1^2 + C_2R_1R_2 - C_2R_1^2b \\ C_1R_2^2b \end{pmatrix} \quad (9)$$

where the second step is due to imposing the normalization condition:

$$R_1a + R_2b = R_1 + R_2. \quad (10)$$

This constraint allows us to reduce the number of variables from 2 to 1 and to derive a close solution.

Finally, normalizing accordingly to equation 3 we have

$$F^{(n+1)} = \begin{pmatrix} \frac{C_1bR_2^2 + C_2R_2^2b + C_2R_1^2 + C_2R_1R_2 - C_2R_1^2b}{C_1bR_2^2 + C_2bR_2(R_2 - R_1) + C_2R_2R_1} \\ \frac{C_1bR_2^2}{C_1bR_2^2 + C_2bR_2(R_2 - R_1) + C_2R_2R_1} \end{pmatrix}. \quad (11)$$

Let's focus only at the second component of the Fitness vector, which in the following we will call $F_2^{(n)}$, and which is referred to the fitnesses of the R_1 lowest rows. Comparing 6 and 11 we see that, in one algorithmic step, we had

$$b \rightarrow \frac{C_1bR_2^2}{C_1bR_2^2 + C_2bR_2(R_2 - R_1) + C_2R_2R_1} \equiv \frac{b}{A_1b + A_2}, \quad (12)$$

where

$$\begin{aligned} A_1 &= 1 + \frac{C_2(R_2 - R_1)}{C_1 R_2} \\ A_2 &= \frac{C_2 R_1}{C_1 R_2}, \end{aligned} \quad (13)$$

from which the closed form solution, by induction, is trivial

$$F_2^{(n)} = \frac{1}{A_1 \sum_{i=0}^{n-1} A_2^i + A_2^n}. \quad (14)$$

The solution for the other component can be obtained from the normalization condition given by Eq. 10

2.3 Convergence

How does $F_2^{(n)}$ behave when n goes to infinity? It trivially depends on A_2 :

- if $A_2 > 1$, F_2 converges to 0 exponentially fast while n goes to infinity;
- if $A_2 = 1$, $F_2^{(n)} = 1/(nA_1 + 1)$ and therefore converges to 0 as fast as n^{-1} ;
- if $A_2 < 1$, F_2 converges exponentially to a non zero number, $(1 - A_2)/A_1$.

From 13 we notice now that A_2 has also a particular geometric meaning: it is the ratio of the areas of the bottom right block (block 4) and the top left block (block 2). Remembering our definitions of the blocks, A_2 is therefore greater than 1, and the second fitness component $F_2^{(n)}$ converges to 0, if the *external area* is bigger than the *internal area*. Or, visually, if the belly of the non-zeros area is inward.

$A_2 = 1$, when A_2 is defined through equation 13, defines the diagonal line, or its obvious geometric extension to rectangular matrices,

$$R_1 = (R_1 + R_2) \frac{C_1}{C_1 + C_2}. \quad (15)$$

It is worth noting that if A_2 is close, but different, to 1, the exponential behavior of the fitness can be very slow. In particular it is possible to see that, for $A_2 > 1$, $F_2 \approx \frac{A_2 - 1}{A_1} e^{-(A_2 - 1)n}$ for A_2 near enough to 1 and n large enough. This define a characteristic time $n^* = 1/(A_2 - 1)$ of the exponential that can be very large when A_2 approaches 1. Similarly for $A_2 < 1$, $F_2 \approx \frac{1 - A_2}{A_1} (1 + e^{-n(1 - A_2)})$, again in the case of A_2 near enough to 1 and n large enough.

2.4 Heterogeneous Density

Interestingly the result does not change if the densities of 1s in the blocks 1, 2 and 3 are different than 1, if the block 4 is still composed only of 0s and the block 2 has a positive density¹

Only the shape matters, and not the density. In particular doing the same computations with densities different from 1 (and potentially heterogeneous) we find that:

- equation 14 does not change,
- A_2 does not change,
- A_1 does depend on the densities of the blocks 1 and 3, but not from the density of block 2 if it is not 0 (see ahead for the case in which the density of block 3 is zero).

The specific value of the density of 1s and 0s in the internal area is therefore completely irrelevant, while the density inside the frontier only determines the specific value of the convergence point (if there is convergence to a number greater than 0).

2.4.1 Zeros outside the frontier

The case in which the density in the internal area is 0 is a peculiar case, since in this case it is not possible to really define an inside and an outside. In fact it is possible to rearrange the rows and columns to switch inside and outside. In this case it is still possible² also to define A_2 . If A_2 is different than 1 the fitness of one of the two sets of countries is always converging to zero. In the case in which A_2 is equal 1 instead, any value of fitness is a stationary point of the algorithm and therefore the fitnesses produced by the algorithm will be equal to the starting conditions.

This can be said in a different way. In this case the blocks 1 and 3 are not connected through a product common among countries belonging different blocks. Therefore the algorithm in this case does not have any direct information to compare fitnesses and complexities of countries and products in different blocks. To compare the two blocks it is therefore only possible to use their densities and shape. However the density turns out to be irrelevant, since the more products in a block the more is diluted their influence on the Fitnesses of the countries in their block: as it is possible to compute, this cancel out the additional effect of having more or less products. Only the relative shape among the blocks, measured by A_2 , is left to compare the two blocks. If A_2 is equal 1, and therefore also the shape does not help to compare the blocks, any fitness and complexity is consistent. If A_2 is different than 1, it will determine the convergence to 0 of the fitnesses and complexities of one of the two blocks.

¹this condition is obviously needed to clearly define an internal area and an external area.

²Of course, in the case in which both block 2 and 4 are composed only of zeros, it is possible. to rearrange rows and columns to switch the blocks, i.e. there are two different ordered M , causing A_2 to go to A_2^{-1} . Therefore A_2 is defined but for a possible multiplicative inversion.

2.5 Ansatz

The ansatz that will be then corroborated by the numerical investigation in the next sections is therefore simple, at least using loose definitions:

Ansatz 1 *If the belly of the matrix is outward all the fitnesses and complexities converge to numbers greater than 0. If the belly is inward, some of the fitnesses will converge to 0.*

While in the case defined in 2.1 the conversational definition of “inward belly” - corresponding to $A_2 > 1$ - is unambiguous, for the general cases that we will investigate in the sections 3 and 4 there are many possible definitions.

Defining

- *ordered* matrix, as the matrix M after rearranging the columns and rows such that all the countries are ranked accordingly to their fitnesses and all the products are ranked accordingly to their complexities³;
- *diagonal* line, as the line, in the previously defined matrix, going from the least fit country and least complex product to the most fit country and most complex product; if the matrix is squared, this definition is the usual definition of diagonal of the matrix; if the matrix is rectangular, this definition is the trivial geometric extension of the previous one.
- the *external* area of the matrix is the joint set of zeros in the ordered matrix including the corner corresponding to the 0 for the lowest fitness and highest complexity product;

we will see that a proper definition, with relative guess, is

Ansatz 2 *A matrix M have all the fitness and complexity different from zero if, after ordering said matrix, the diagonal line does not pass through the external area.*

It is worth to note that when a country’s fitness converges to zero, from equation 2, we know that all the exported products’ complexity will converge to zero too. At the same time, the effect on the fitnesses of countries which export a product whose complexity converges to zero, from equation 1, is negligible. Therefore, we can consider that, but for multiplicative common factors coming from equations 3 and 4, removing from the matrix M a line corresponding to a country converging to zero fitness and all the columns corresponding to its exported products, the output will be not affected.

It is therefore reasonable to make a further guess:

Ansatz 3 *If the diagonal line does pass through the external part of a ordered matrix M , some countries and products will converge to zero. If we progressively remove them from the analysis, defining new ordered matrices M , we will have a convergence to finite values of fitnesses and complexities for the remaining countries and products for which the the new M have a diagonal line not passing through the external area.*

These ansatz implicitly define the *crossing country* of a matrix as the lowest fitness country that converges to a nonzero fitness. In other words, it is the first country that does not need to be removed in the removal process described above, defining the largest matrix such that all its countries have nonzero fitness. Note that in general searching instead from the top countries does not give the proper result in a situation in which the diagonal cross multiple times the external area. If one starts the check from the top countries and keeps adding countries with lower fitnesses, the crossing country is not the first one met. In this case of multiple crossing, the crossing country is the first country met in the process starting from the bottom, as in ansatz 3.

The guesses 2 and 3 will be tested numerically in the sections 3 and 4.

3 A numerical investigation

In this section we present the results of our numerical simulations, in which we study some peculiar behaviors of sample matrices which we did not treat analytically. In particular, by means of synthetic examples of possible M matrices we study the cases in which more than two blocks are present, the importance of producing rare products and we investigate the ansatz proposed in the previous section. The reader will notice that the matrices investigated in the first two subsections correspond to the case $A_2 = 1$ discussed above.

³while in same cases, as we have seen and we will see, the fitness could converge to 0 for many countries, for any finite number of iteration the fitness values will be greater than 0. Moreover, since the convergence speed after some iterations is constant for each country and products, there will be a number of iterations such that, for any following iteration, the ranking of fitnesses and complexities will be constant.

3.1 More than two blocks

In the case of more of two blocks it is not possible to write the components of $F^{(n)}$ in the form given by Eq.6, because one would have more than one unknown variable. However, what one can do is progressively reduce the given matrix to submatrices, each one made of two blocks. In the following example we present a block diagonal matrix:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

We have found that if off block diagonal elements are absent (that is, both the internal and the external areas are empty) all countries will remain in the respective initial conditions. As we expected from the two block analysis presented in Sec.2.4.1, this result is stable even if density is lower: only the relative dimensions of the blocks counts.

3.2 The importance of oligopolies

The presence of common products, in the very spirit of [Tacchella et al., 2012], does not add much information other than the fact that those countries which produce only very common products can not be very high in the fitness ranking. As a consequence, we can expect *monopolies* of already diversified countries to be relevant. To investigate the consequences of having common ubiquitous products we take into account a synthetic situation in which three countries produce only very common products, two countries have a *duopoly* and one country, which has the same diversification of these two countries, has a monopoly. The results are the following:

$$\begin{matrix} c \\ n^{-1} \\ n^{-1} \\ n^{-2} \\ n^{-2} \\ n^{-2} \end{matrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

On the left side of the matrix we show the speed of the fitness decays with the iteration n , with c meaning convergence to a finite value of fitness⁴. For example, the notation n^α means that the fitness $F_c^{(n)}$ of the country c corresponding to that row is converging towards zero with a power law of the number n of the algorithm iterations, that is, $F_c^{(n)} \sim n^\alpha$ for a sufficiently larger n . As one can notice, the presence of a monopoly makes the first country the only one to converge to a nonzero value of F , while the advantage given by the extra products makes the second and third country perform better with respect to the last three countries. In any case, this advantage is not striking, because we are still in the case in which $A_2 = 1$, so the difference is only in the exponent of the decay. We stress that the presence of a different decay means that there exist a number N so that from iteration $n = N$ the ranking will stay constant, and so the ranking will be given by the decay exponent.

A further, self-illustrative situation of the consequences of having common products is presented below.

$$\begin{matrix} c \\ n^{-1} \\ n^{-2} \\ n^{-3} \end{matrix} \begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Let us now consider the following matrix:

$$\begin{matrix} c \\ c \\ n^{-0.6} \\ n^{-1} \\ n^{-1} \\ n^{-1} \end{matrix} \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

here we have three blocks of nations, with one monopoly (country 3) and two duopolies (countries 1 and 2). On the basis of the previous considerations, we know that the 3x3 submatrix constituted by countries 1,2 and 3 and products 4, 5 and 6 is degenerate, in the sense that every initial condition $F^{(0)}$ is stable under the iterations (also the whole matrix would be degenerate if country 1 would not have produced product 2). However, in this case, the presence of an external product changes the situation, giving an advantage to the duopoly, which converges to a finite value of the fitness and

⁴In this section and in the following we will not report the complexity decays, which follow trivially from the ones of the countries.

makes the monopoly tend to zero, even if with a lower exponent (0.6 instead of 1). By means of extensive numerical simulation we have found that the specific value of the decay is given by the relative size of the worst block (the one which decays with n^{-1}) with respect with the total size of the externally connected block, that is, the sum of the sizes of the blocks which are connected by the external 1. In other words, the decay is given by the ratio between the worst block and the sum of the sizes of all the blocks but the one whose decay we are computing. In the situation presented above, the worst block has size 3 and the converging block has size 2, from which we have $3/(3+2) = 0.6$. To make another example we can imagine a different case, in which we give an advantage to the monopoly with respect to the block of size 3; in this case, the duopoly will decay with an exponent equal to $3/4$.

Another interesting phenomenon is evident. Let us consider the set of countries 4, 5 and 6. Even if the second product is owned only by countries 4 and 5, also country 6 decays in an equally faster way. This is due to the fact that the last block is *connected*, that is, there exist a path of products that connects all the countries in the block, and the absence of monopolies in the block.

3.3 Exponential decays

Until now we have seen that different power law decays come out from, in general, N-polistic competitions whose symmetry is broken by the presence of products external with respect to the competitors. In the case in which one or more N-poly has a further N-poly which is not shared, the decay of the competitor will be *exponential*. Here it is an example:

$$\begin{matrix} c \\ e^{-n} \\ e^{-n} \\ e^{-n} \end{matrix} \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

In this case the first country has two monopolies; country 3 one monopoly and countries 2 and 4 no monopolies at all. All countries tend to zero exponentially faster, but the first one. This is a simple example of an inward bellied matrix, that is, a matrix such that the diagonal crosses the external empty part. Because in this case the crossing country is the first one, all the other countries will converge to zero exponentially, as stated in the ansatz 3. This phenomenon will be investigated in detail the next subsection.

3.4 Numerical verification of the ansatz

In Sec.2.5 we suggested a link between the convergence properties of countries and the shape of the belly of the ordered matrix. In this section we verify the given ansatz by means of simple matrices, whose behavior is nevertheless analogous to the one we will find in real world applications. Let us start with a set of 5x5 square matrices. We point out that we use square matrices only because the diagonal and, as a consequence, the shape of the belly is evident by eye. As we will see in the following, these results can be applied also to rectangular matrices by means of the generalized definition of diagonal discussed above.

$$\begin{matrix} c \\ n^{-1} \\ n^{-2} \\ n^{-3} \\ n^{-4} \end{matrix} \begin{matrix} A \\ \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix} \quad \begin{matrix} c \\ e^{-n} \\ e^{-n} \\ e^{-n} \\ e^{-n} \end{matrix} \begin{matrix} B \\ \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix} \quad \begin{matrix} c \\ c \\ c \\ c \\ c \end{matrix} \begin{matrix} C \\ \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix} \end{matrix} \quad \begin{matrix} c \\ n^{-1} \\ n^{-1} \\ n^{-1} \\ n^{-1} \end{matrix} \begin{matrix} D \\ \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

Matrix *A* represents a borderline case, in which the same lines of reasoning adopted in Sec.3.2 can be applied. All countries but one are converging to zero, preserving a well defined ranking.

Matrix *B* shows a clear inward belly. This means that, according to our ansatz, in order to know which countries will converge to a nonzero value of fitness we have to eliminate countries starting from the ones with a lowest ranking until we find a submatrix with an outward belly. In this case this is not possible, in the sense that the only fully convergent submatrix would be the trivial 1x1 matrix containing only country 1 and product 5. As a consequence, only the country 1 (which, in this case, is the trivial crossing country) and product 5 converge to a nonzero value of fitness and complexity, while the other countries and products decay exponentially.

On the contrary, matrix *C* has a clear outward belly: all countries have products beyond the diagonal and so converge to nonzero values of fitness. An interesting situation is shown in matrix *D*, which is equal to *C* but for one product. By removing product 4 from country 2 is it possible to make all countries but one converge to a zero value of fitness. In fact, in this case the diagonal crosses the external part of the matrix in correspondence with an high ranking country. This situation highlights the importance, in general, of a careful data sanitation when empirical *M* matrices are built, not only with respect to low fitness countries, but also for high fitness ones. In any case, we point out that when one passes from matrix *C* to *D* the ranking is preserved.

In order to show a case in which the *M* matrices are rectangular and the crossing country is nor in the top, neither in the lowest position in the matrix, but somewhere in the middle of it, we consider the three following matrices:

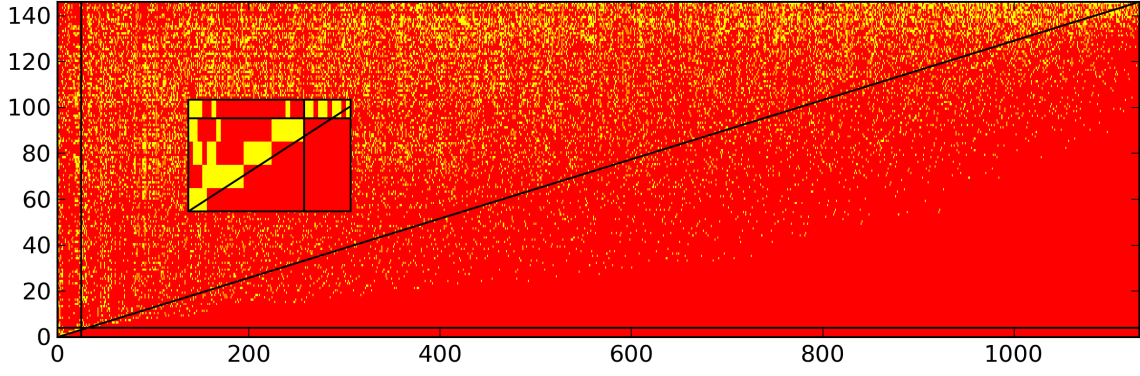


Figure 1: Ordered M matrix for the BACI dataset, year 1995. The diagonal is consistently above the external part. However, there is an exception for the low fit countries, zoomed in the picture inset. The horizontal and vertical lines represent the number of countries and products that converge to zero.

$$\begin{array}{ccc}
 E & F & G \\
 c \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} & c \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ n^{-1} \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} & c \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ e^{-n} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}
 \end{array}$$

While matrix E shows full convergence of its countries, it is enough to remove one product to make the external area large enough to be crossed by the diagonal and, as a consequence, to make two countries converge to zero (matrix F). The removal of one more product turns the convergence to zero exponential with the number of iterations. We stress that, even if the number and the category of products of country 4 is the same for all the three situations, its value of fitness changes. However, this fact is not paradoxical as long as the values of the matrices are calculated in a correlated way, by considering, for example, the Revealed Comparative Advantage.

4 Real cases

In this section we study how this variety of convergence behaviors affects real M matrices. Typical M matrices are bigger than the ones seen in chapter 3 and present specific characteristics of nestedness and density; nevertheless, our convergence ansatz turn out to be relevant for real M matrices.

4.1 UN COMTRADE dataset, 1995-2010

The first database we consider is based on the import-export flows of products among countries collected by the United Nations and processed by BACI [Gaulier and Zignago, 2010] and covers the years from 1995 to 2010. The number of products is fixed through the years and equal to 1131, classified accordingly to the 4 digits Harmonized System 2007, while the number of countries slightly varies between 146 and 148. A data sanitation procedure has been performed, and the resulting values of the M matrices are assigned by means of a threshold on the Revealed Comparative Advantage (RCA) as introduced by Balassa [Balassa, 1965]. In all the years the matrix is outward bellied, accordingly to the definition implicit in 2. In other words, the diagonal does not cross the external (i.e., empty) part of the matrix, and so the algorithm converges for most countries to a non zero value of fitness. While the general behavior is clear, some exceptions are there for the least fit countries, as it is visible in figure 1. In the figure we show a pictorial representation of the ordered M matrix for the year 1995. The yellow (red) dots mean that the country corresponding to the given row is (not) exporting (in the RCA sense) the product corresponding to the given column, that is, the matrix element is equal to 1 (or zero, respectively). The horizontal line corresponds to the *crossing country* defined in 2.5, and the vertical one to the product with higher complexity exported by it. The part of the matrix which converges to nonzero values of fitness and complexity is defined by these lines (in particular, the top right portion of the original matrix). The dashed line is the diagonal of the fully convergent matrix. Obviously, after the removal process the matrix is outward bellied in all its parts.

It is interesting to note that the number of countries and products that need to be removed to have a matrix converging to non-zero values is reduced for the later years. This behavior is general, for all the datasets checked. The causes of this phenomenon will be discussed in the conclusions in section 6.

4.2 UN COMTRADE dataset, 1963-2000

A second dataset we analyze is the one collected by the United Nations and processed by Feenstra et al. [Feenstra et al., 2005]. It contains the import-export flows of 538 products, classified accordingly the Sitc rev2 products classification, for the

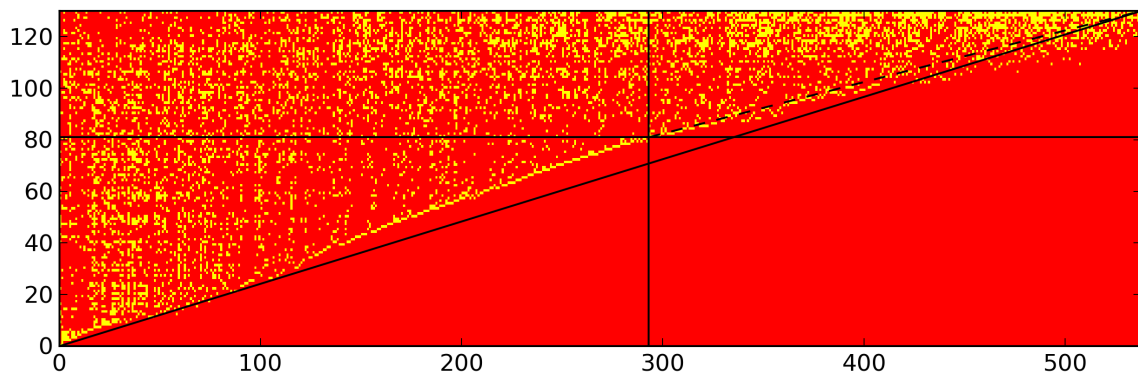


Figure 2: Ordered M matrix for the 1965-2000 dataset, year 1965. The matrix frontier is consistently above the diagonal line. The horizontal and vertical lines represent the number of countries and products that converges to zero. The dashed line shows the diagonal of the new block remaining after removing those countries and products: the diagonal of the new M matrix now does not cross the external part of the matrix.

years 1963-2000. The number of countries ranges from 131 to 154. The dataset is particularly interesting because of the long time span of data presented in a consistent form. In this dataset the convergence ansatz proposed is more relevant, since big parts of it is consistently above the diagonal. An example, that highlight also an application of ansatz 3, is presented in figure 2.

The example also highlights a peculiar graphical feature of the ordered matrix. Indeed, for the countries and products that are converging to zero it is possible to observe how the frontier between the internal and external part is smooth and continuous, while a more rough behavior can be observed for the areas converging to positive numbers. This behavior is due to the fact that, since the fitnesses and complexities of the countries and products below the crossing country are converging exponentially to zero, for any given finite iteration the ratio between two consecutive fitnesses and complexities is also exponentially growing. Therefore the fitnesses are mostly determined by the highest complexity product and, similarly, the complexities are given by the lowest fitness country. As a consequence there is a clear correspondence between best products and worst countries, determining the smooth frontier that it is clearly visible in figure 2. At the opposite, for the countries above the crossing country, their relative fitnesses converges to fixed ratios. Therefore the fitnesses of the countries are determined by all the products not converging to zero, causing their ranking to be not uniquely determined by their best product. A similar line of reasoning can be used for the products. Therefore after the crossing country there is not a clear frontier.

A similar behavior can be observed in all the matrices that we studied. Even in the case of figure 1 it is possible to recognize the phenomenon for the last four countries (zoomed in the inset).

4.3 Patents

The convergence to zero of a sizable set of countries is very evident for matrices particularly empty, like the one produce from the patents dataset. It is a dataset organized by OECD starting from data produced by the European Patent Office⁵ joining countries and technological sectors if a firm in a country has patents granted in that sector. It is then manipulated in the usual way, through consideration of comparative advantage to remove the size of the countries from the argument, generating a matrix that has a 1 or 0 if it produce more patents than expected in a particular technological sector. Since to patent a discovery in a sector the country has to be on the frontier of the technological progress in that sector, the matrix comes out with an inward belly, as it is visible in figure 3.

As for guess 3, a possible situation leading to a global convergence to zero is for a single country being diversified in too many sectors in which no other countries is interested in. In this case we can have all the other countries' fitnesses converging to zero. An example is in figure 4.

5 Convergence in ranking: a practical note

A practical application of our results is that the usual convergence conditions are not suitable for this algorithm. In particular, the condition $|F^{(n)} - F^{(n-1)}| < \epsilon$ does not imply that the single components of the fitness vector stop to decrease and, in general, the resulting fitnesses and complexities could depend on ϵ . This could be particularly relevant if one is interested in non linear expressions of fitness or complexities (e.g. the logarithm of the countries' fitness), also given the fact that convergence rates can be very slow.

In principle however, one may be interested only in the rankings. Obviously, different criterions can be assessed for the ranking convergence. Indeed the problem is not trivial because of the very slow convergence of the fitness described in section 2.3. While the exact power law behavior $F_c^{(n)} = F_c^{(n_0)} \left(\frac{n}{n_0}\right)^{\alpha_c}$ described in section 3.2 is less likely to be found

⁵For more informations, see the OECD Patent Statistics Manual available at <http://browse.oecdbookshop.org/oecd/pdfs/free/9209021e.pdf>

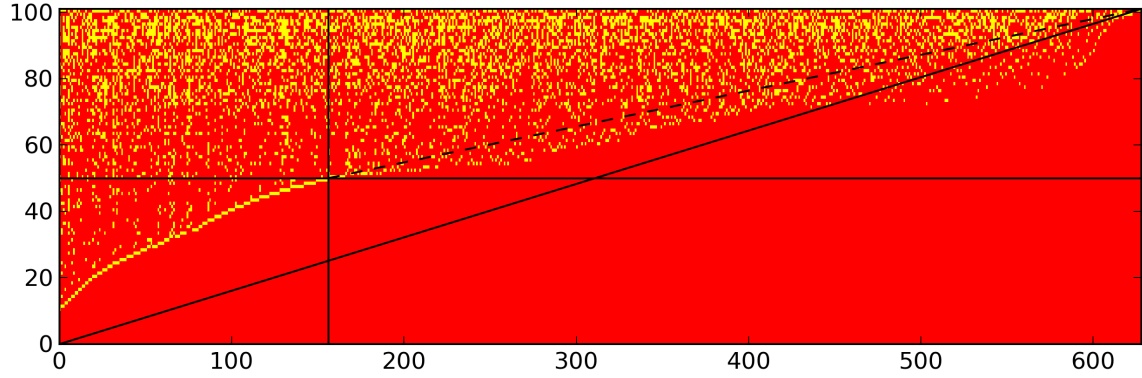


Figure 3: Ordered M matrix for the patents dataset, year 2002. The external part is very big, so diagonal line. crosses it and a consistent fraction of countries converges to zero. The dashed line shows the diagonal of the new block remaining after removing those countries and products.

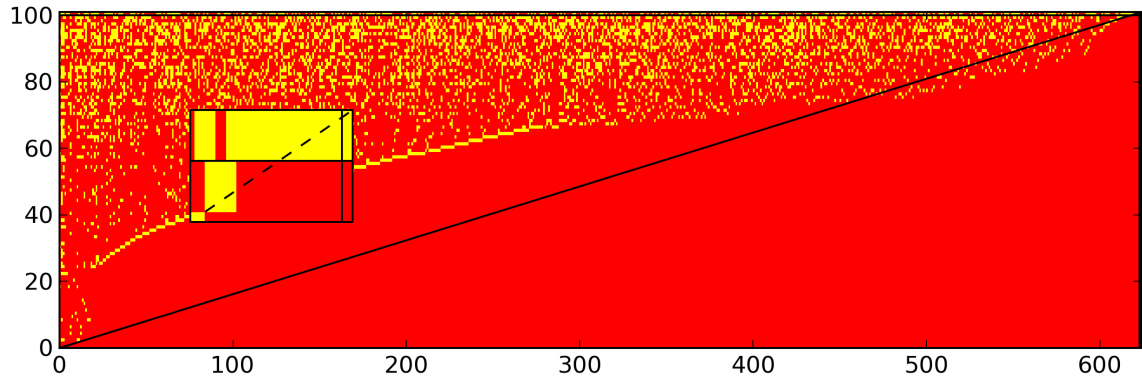


Figure 4: Ordered M matrix for the patents dataset, year 2000. The diagonal line is consistently inside the external area. More importantly, this is true even for just the top two countries, as it is visible in the zoomed in the picture inset. As a consequence, all the countries and products converges to zero but the first one, as shown by the horizontal and vertical lines. The dashed line shows the diagonal of the block that would be left removing all countries and products but the first two: the diagonal would still cross the external part. Therefore the only remaining non-zero-fitness country is the first one, and the only remaining non-zero-complexity products are the ones produced only by it.

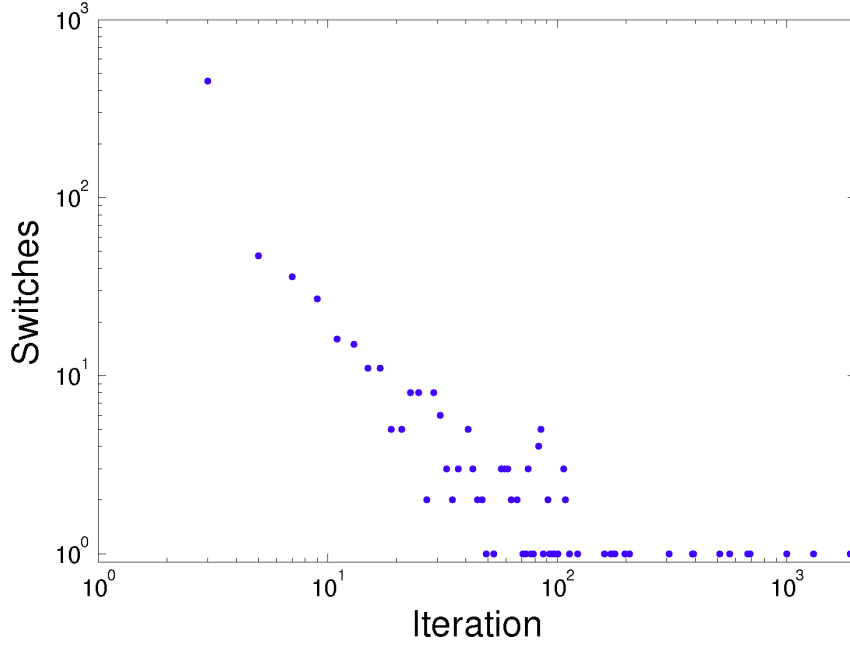


Figure 5: Number of crossings from one iteration to the next one, as a function of the total number of iterations. From the fat tail behavior shown it is clear that crossings can occur even after thousands of iterations. Data from UN COMTRADE dataset, year 1990.

in larger real matrices, the exponential behavior $F_c^{(n)} = F_c^{(n_0)} e^{(n-n_0)/n^*}$ can have a very large characteristic time n^* , as noted in section 2.3, so that can be considered approximately power law for a large number of iterations. Therefore the ranking can change even after many iterations of the algorithm. It is better to see the issue in a practical case.

In figure 5, we show how the ranking keeps changing in a real case even after thousands of iteration steps. Indeed we can see how, in the context of a power law approximation, the approximated exponent α_c can be very small for many iterations. Note that a value of α_c of 0.1 means that one must wait 10^{10} iterations for the fitness to reduce of one order of magnitude. It is therefore clear that when two countries' fitness are decaying with very similar α_c , to switch ranks could take any amount of time. We can see an example for two such countries, related to the previous real case, in figure 6.

While it is impossible to assure that there will be no crossing in the next iterations, one can estimate a lower bound to the number of iterations still needed for the next change in ranking. This is computed as follows:

- We estimate the growth rates α_c for any country c , by the formula:

$$\widehat{\alpha}_c = \frac{\log(F_c^{(n)}) - \log(F_c^{(n-\delta)})}{\log(n) - \log(n-\delta)}$$

where δ is a suitable delay⁶.

- We order the countries accordingly to their fitness to have $F_c^{(n)} > F_{c+1}^{(n)}$.
- For any c we estimate the crossing iteration CI_c between the countries c and $c+1$ accordingly to

$$\widehat{CI}_c = n \left(\frac{F_c^{(n)}}{F_{c+1}^{(n)}} \right)^{1/(\widehat{\alpha}_{c+1} - \widehat{\alpha}_c)}.$$

Of course the quantity \widehat{CI}_c is a reasonable estimate of an actual possibility of a rank change only if $\widehat{CI}_c > n$, i.e. $\widehat{\alpha}_c > \widehat{\alpha}_{c+1}$. In the other case, when $\widehat{\alpha}_c < \widehat{\alpha}_{c+1}$, the two fitnesses are indeed diverging.

- Among the valid crossing iterations $\widehat{CI}_c > n$, we define the Minimum Crossing Iteration as

$$MCI^{(n)} = \min_c(\widehat{CI}_c) \quad (16)$$

The value of $MCI^{(n)}$ is therefore, for any n , a lower bound for the iteration when the next rank change will occur. Its value can be used in order to assess when to stop the algorithm. This value is indeed a lower bound since the real exponential decay is always faster than the estimated power law. However, since the characteristic time of the exponential is very slow, the lower bound is very close to the real value. This can be seen for example in the real case

⁶in the following we will use $\delta = 2$.

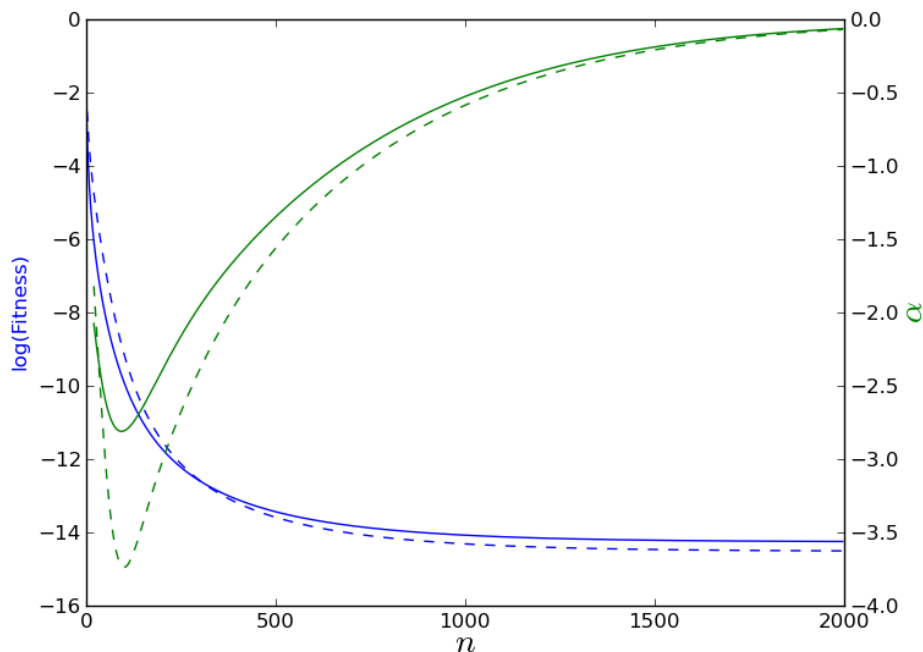


Figure 6: Comparative analysis of the behavior of the fitness and estimated growth rates for two sample countries near the crossing country. In blue the logarithm of Fitness, accordingly to the left y-axis, in green the growth rates. The two countries are identified by a solid or dashed lines. It is possible to see that both the growth rates and the countries' fitness cross after many iterations. In particular the ranking switches at around 500 iterations. It is interesting to note that the switch in rankings of the other countries completely changes the behavior of the two countries: indeed, until around 80 iterations the two fitnesses seemed to exponentially converge to zero. However, after that point, the converging behavior emerges, as it is visible looking at the growth rate evolution. Data from UN COMTRADE dataset, year 1990.

behavior of $MCI^{(n)}$ shown in figure 7. Comparing figure 7 and figure 5 gives an insight on the algorithm. In particular one can precisely match the crossing times and the predicted minimum crossing times. After each crossing $MCI^{(n)}$ jumps to predict the next crossing. Notice that the observed plateaux of $MCI^{(n)}$ show the precision of the power law approximation: the forecast would not change at all until the next crossing only if the power law behavior was exact.

As a practical note, if one is interested only in the ranking, one can set a threshold MCI to the maximum number of iterations that he is a priori willing to take. Of course it is also possible to define an equivalent measure for the ranking of products, that would give a different stopping time for the algorithm.

In the practical spirit of this section, it is important to stress that the number of iterations could get very high to achieve a satisfying value of MCI . In this case it is possible that, while some countries (or products) are still slowly decaying, others are already well below the precision of the machine used for computations. Some technical attention is therefore required to avoid that the algorithm fails to compute, since the Fitness-Complexity algorithm 1-4, is ill defined for zero values of fitness.

6 Conclusions and discussion

This paper shows that the intrinsic non linear structure of the algorithm proposed by [Tacchella et al., 2012] has highly non trivial consequences on its convergence properties. In particular, we have linked the shape of the ordered matrix on which the algorithm is based to the possible presence of countries whose fitness tends to zero. Obviously, also all the complexities of their exported products will converge to zero. In synthetic but general cases it is possible to show analytically that the presence of a large empty part of the matrix makes some countries converge to zero. In particular, the condition to check is if the matrix diagonal crosses or not the empty part of the matrix. We study numerically some simple cases which confirm our results also when the analytical approach was not possible and in some real world cases, such as two different country-product databases and a country-patent database. We extend our results to generalized versions of the algorithm, finding similar results. We also give a practical recipe to decide when to stop the algorithm if interested in a stable ranking.

A striking result of our analysis is that a large variety of situations is present. For example, if one considers the UN COMTRADE dataset 1995-2010, almost all countries have a finite fitness, while in many years of the UN COMTRADE dataset 1963-2000 more than a half of the countries converges to a zero value of fitness. On the other hand, in some patents datasets all the countries fitnesses but one tend to zero.

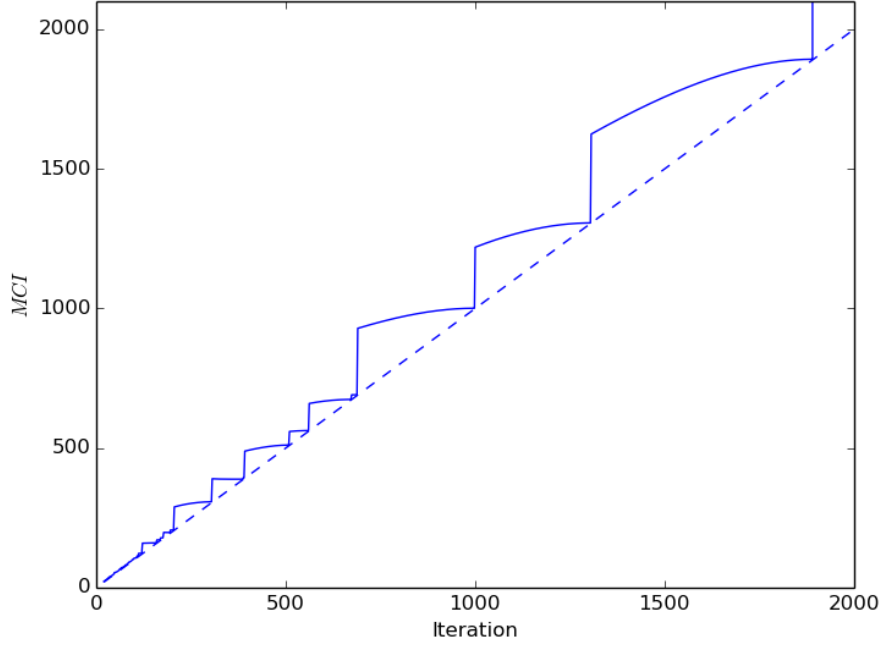


Figure 7: Evolution of the Minimum Crossing Iteration, as defined in the text, as a function of the iteration. The expected number of iterations one should wait to observe another change in rank appears to be a step like function jumping after each crossing (see figure 5 for a comparison). The increasing behavior confirms the lower bound property of the measure. After the last crossing at the iteration 1892, the next predicted change in ranking is $3.7 \cdot 10^6$ iterations. As a consequence, if I am interested only in the country ranking and I am not willing to run the algorithm for millions of iterations, it is reasonable to stop the algorithm here.

References

- [Tacchella et al., 2012] Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., and Pietronero, L. (2012). A new metrics for countries’ fitness and products’ complexity. *Sci. Rep.*, (2):723.
- [Feenstra et al., 2005] Feenstra, R.C., Lipsey, R.E., Deng, H., Ma, A.C., and Mo, H. (2005). World trade flows: 1962-2000. Technical report, National Bureau of Economic Research.
- [Gaulier and Zignago, 2010] Gaulier, G., and Zignago, S., (2010). <http://www.cepii.fr/anglaisgraph/workpap/pdf/2010/wp2010-23.pdf> Baci: International trade database at the product-level. Centre d’Etudes Prospectives et d’Informations Internationales.
- [Balassa, 1965] Balassa, B. (1965) Trade liberalisation and “revealed” comparative advantage. *The Manchester School* 33: 99–123.
- [Cristelli et al., 2013] Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G., and Pietronero, L. (2013). Measuring the intangibles: A metrics for the economic complexity of countries and products. *PLoS ONE*, (8):e70726.
- [Barro and Sala-i Martin, 2004] Barro, R. J. and Sala-i Martin, X. (2004). *Economic Growth*. MIT Press.
- [Hidalgo and Hausmann, 2009] Hidalgo, C. and Hausmann, R. (2009). The building blocks of economic complexity. *P. Natl. Acad. Sci. USA*, 106:10570–10575.
- [Ricardo, 1817] Ricardo, D. (1817). *On the Principles of Political Economy and Taxation*. John Murray.

A A generalization to a wider class of algorithms

Many possible variations of the algorithm could be implemented changing equations 1-4 with the purpose of not having zero solutions. The most obvious one is probably a variation of the harmonic mean in eq 2 with a different elasticity coefficient, such as

$$\tilde{Q}_p^{(n)} = \left(\sum_c M_{cp} \left(F_c^{(n-1)} \right)^{\frac{1}{\gamma}} \right)^{\gamma}. \quad (17)$$

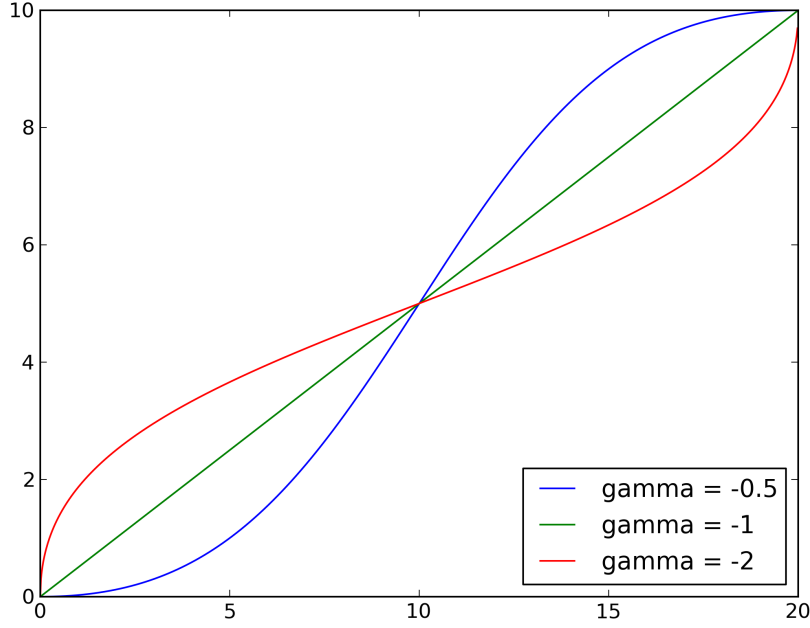


Figure 8: Line $A_2 = 1$ for different values of γ , for a matrix M 10×20 . The algorithm described in equations 1-17 converges to all values different from 0 if the frontier is all below the line.

For $\gamma = -1$ equation 17 goes back in equation 2, as in the original algorithm. For $\gamma = 1$ equation 17 becomes a sum, symmetrically to equation 1. This is in strong opposition to what the algorithm is meant to describe, since it would mean that the more countries produce a product the higher its complexity. Therefore we will limit ourselves to discuss the cases $\gamma < 0$.

Computations similar to the ones in section 2 can be done in this case, obtaining similar conditions. Equation 12 is still valid, even if only in the limit $b \ll a$, and therefore after many iterations⁷. However the parameter A_2 in this case is equal to

$$A_2 = \frac{C_2}{C_1} \left(\frac{R_2}{R_1} \right)^\gamma. \quad (18)$$

Also ansatz 2 and 3 change accordingly. While $A_2 = 1$ with A_2 defined in equation 13 identified the diagonal line, when we define A_2 through equation 18 we have that imposing $A_2 = 1$ we identify the curve

$$R_1 = (R_1 + R_2) \frac{C_2^{1/\gamma}}{C_1^{1/\gamma} + C_2^{1/\gamma}}. \quad (19)$$

Examples for different values of γ are presented in figure 8. From the figure it is clear the trade-off of changing the value of γ to avoid values of the fitness converging to 0: to loosen the condition for high fitness countries, one makes it stricter for low fitness countries, and viceversa. Moreover it is also clear that changing γ does not help if the frontier is above the diagonal in the middle.

⁷the equation 12 is exactly valid for any b , and not only in the limit $b \ll a$, for the case in which the density of block 2 is 0. In that case, changing 13 with 18, everything works exactly as in the section 2.4.1